DOCUMENT RESUME

ED 342 782 TM 017 909

| | |
|---|---|
| AUTHOR | Bossone, Richard M., Ed.; Polishook, Irwin H., Ed. |
| TITLE | Educational Assessment. Proceedings: Conference of the University/Urban Schools National Task Force (13th, Atlanta, Georgia, November 8-9, 1991). |
| INSTITUTION | City Univ. of New York, N.Y. Graduate School and Univ. Center. |
| PUB DATE | 92 |
| NOTE | 92p. |
| PUB TYPE | Collected Works - Conference Proceedings (021) |
| EDRS PRICE | MF01 Plus Postage. PC Not Available from EDRS. |
| DESCRIPTORS | Conference Papers; Curriculum Development; *Educational Assessment; Elementary Secondary Education; Evaluation Methods; *Instructional Improvement; Mathematics Instruction; *Policy Formation; Program Evaluation; Reading Instruction; Science Instruction; *State Programs; Student Evaluation; Testing Problems; *Testing Programs; Workshops; Writing Instruction |
| IDENTIFIERS | *Authentic Assessment; National Education Goals 1990; Performance Based Evaluation |

ABSTRACT

At this conference on educational assessment, there was consensus among participants on the importance of the use of assessment to improve instruction, enhance the curriculum, and allow policy makers to measure achievement. There was less consensus about the best forms of assessment to use. The following conference papers are provided in this document: (1) "The National Education Goals Report" (P. D. Forgione, Jr.); (2) "Assessing the Business Roundtable's 1990s Education Initiatives" (M. T. Nettles); (3) "National Assessment for Educational Progress" (A. E. Lapointe); (4) "Georgia State Assessment Program" (S. Bernknopf); (5) "Los Angeles City School Assessment Program" (F. I. Stevens); (6) "Philadelphia School District Assessment Program" (R. C. Altman); (7) "Assessment in Mathematics Instruction" (T. J. Cooney); (8) "Assessment in Science Instruction" (A. B. Champagne); (9) "Assessment in Reading and Writing Instruction" (S. W. Valencia); and (10) "Summation" (R. Rothman). (SLD)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY
RICHARD M. BOSSONE
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED342782

# PROCEEDINGS:

The Thirteenth Conference of the
University / Urban Schools National Task Force

# Educational Assessment

Edited by Richard M. Bossone
and Irwin H. Polishook

The Graduate School and University Center
of The City University of New York

TM017909



BEST COPY AVAILABLE

# PROCEEDINGS:

The Thirteenth Conference of the
University / Urban Schools National Task Force

# Educational Assessment

November 8 and 9, 1991
Atlanta, Georgia

Edited by Richard M. Bossone
and Irwin H. Polishook

The Graduate School and University Center
of The City University of New York
1992

## PREFACE

When one compares the cycles of alarm about public education, one finds a constant concern with testing. Each generation seems to return to testing issues as a token of education's enduring power to improve individuals and transform the nation. In the past decades, there has been an ongoing evaluation of the efficacy of I.Q. surveys, college placement examinations, and, more recently, severe challenges have been made to uses and abuses of "standardized" testing in all grades of education, public service, business and the professions. Our current engagement with these issues under the signature word "assessment" is nothing more than a contemporary refinement of these long standing debates about the value of testing instruments and their usefulness in validating our academic and social concerns.

What stands revealed in these papers is more than just a recasting of old subjects in new formulations. The eagerly sought "authentic" assessment movement requires agreement on the goals of each segment of a complex universe of national, state, and schoolbased evaluations. Moreover, competing techniques and values are contained within diverse assessments, whether focused on public accountability, rather than individual student achievement, or multiple choice formats, rather than performance based examinations. Without a nationwide effort to reconcile the distinct meaning of these procedures and their contributions to educational reform, each of these assessments promises even greater testing confusion.

At this conference, there was a consensus among the participants about

the importance of making assessment work to improve instruction, enhance the curriculum, and afford education policy makers the opportunity to measure what immense expenditures of public and private monies have achieved over the past decade; but there was less consensus about the best forms of assessment to employ. Nevertheless, the dialogue here is provocative and should serve as catalyst to all who must make decisions regarding educational assessment.

R.M.B.

I.H.P.

TABLE OF CONTENTS


Page

PREFACE . . . . . ii

PARTICIPANTS . . . . . vi

EDUCATIONAL ASSESSMENT: NATIONAL PERSPECTIVES

The National Education Goals Report
Pascal D. Forgione, Jr. . . . . 1

Assessing the Business Roundtable's 1990s Education Initiatives
Michael T. Nettles . . . . . 15

National Assessment for Educational Progress
Archie E. Lapointe . . . . . 19

EDUCATIONAL ASSESSMENT: STATE, CITY, AND DISTRICT PERSPECTIVES

Georgia State Assessment Program
Stanley Bernknopf . . . . . 29

Los Angeles City School Assessment Program
Floraline I. Stevens . . . . . 35

Philadelphia School District Assessment Program
Rita C. Altman . . . . . 43

Page

SPECIFIC INSTRUCTIONAL ASSESSMENT

Assessment in Mathematics Instruction
Thomas J. Cooney . . . . . 51

Assessment in Science Instruction
Audrey B. Champagne . . . . . 59

Assessment in Reading and Writing Instruction
Sheila W. Valencia . . . . . 66

SUMMATION Robert Rothman . . . . . 76

THIRTEENTH CONFERENCE OF THE UNIVERSITY/URBAN SCHOOLS NATIONAL TASK FORCE

CONFERENCE PARTICIPANTS

Chairman

Dr. Richard M. Bossone
University Dean for Instructional Research
Graduate School and University Center
City University of New York

Participants

Stephen R. Aiello
Vice President and General Manager
Cohn & Wolfe

Anne Alexander
Vice President, Educations Programs
AT&T Foundation

Rita C. Altman
Associate Superintendent
School District of Philadelphia

Barbara V. Bell
Principal, Joe Hall Elementary School
Dade County Public Schools, Miami

Mary Anne Bellinger
Vice President
Atlanta Board of Education

Stuart Berger
Superintendent
Wichita Public Schools

Stanley Bernknopf
Director, Division of Assessment
Georgia Department of Education

Wendy L.K. Best
Assistant Manager
BellSouth Foundation

William C. Bonifield
Vice President for Education
Lilly Endowment, Inc.

Alfred Bork
Professor, Department of Information and Computer Science
University of California, Irvine

Robert F. Boruch
University Trustee Professor, Graduate School of Education
University of Pennsylvania

Lisa Brandes
National Education Advisor
Josten Learning Corp., Washington, D.C.

Wayne Carlson
Director, Partnerships and Adopt-A-School Program
Los Angeles Unified School District

Betty Castor
Commissioner of Education
Florida Department of Education

John E. Cawthorne
Senior Research Associate
Center for the Study of Testing, Boston College

Audrey B. Champagne
Professor of Education and Chemistry
University of Albany, SUNY

Clark T. Colvin
Vice President
Cajun Electric Power Co-operative, Inc.

Thomas J. Cooney
Department of Mathematics Education
University of Georgia

Jim Crain
Chairman
Massachusetts State Board of Education

Joseph B. Cumming
Professor Emeritus
West Georgia College

Anne T. Dowling
Director, Corporate Contributions
Philip Morris Companies Inc.

Fritz Edelstein
Senior Fellow
National Alliance of Business

Edith Everett
Vice Chairperson, Board of Trustees
City University of New York

Henry J. Everett
President
Lexington Associates, New York City

Pascal D. Forgione, Jr.
Executive Director
National Education Goals Panel

Lois V. Freeman
Assistant Superintendent, Instructional Services
East Side Union High School District, San Jose

Pearl Gasarch
Secretary, PSC/CUNY
City University of New York

Anthony Golden
Director, PACAT and Professor Psychology
Austin Peay State University

Michael Grady
Senior Research Associate
The Annie E. Casey Foundation

Donald R. Greene
President
The Coca-Cola Foundation

Clement E. Hanrahan
Vice President/Executive Director
The UPS Foundation

Linda N. Hansche
Director, Georgia Assessment Project
Georgia State University

Fred Hechinger
Senior Advisor
Carnegie Corporation of New York

Willie Herenton
Mayor
Memphis, Tennessee

Thomas Y. Hobart, Jr.
President
New York State United Teachers

Lawrence A. Ianni
Chancellor
University of Minnesota, Duluth

Donald G. Jacob
Director, State Government & Civic Affairs
Pfizer Inc.

Arthur Jefferson
Professor, Educational Studies Department
School of Education, University of Michigan

Lyle V. Jones
Director, Thurstone Psychometric Laboratory
University of North Carolina, Chapel Hill

Phyllis Kaminer
President
Instructional Systems, Inc.

Archie E. Lapointe
Executive Director
Center for Assessment of Educational Progress

Richard L. Larson
Professor of English, Lehman College
City University of New York

Leonard Lund
Program Director, Business/Education Research
The Conference Board

Gerald W. Lynch
President, John Jay College of Criminal Justice
City University of New York

Shauna L. MacIntosh
Program Director, Utah Academy of Finance
National Academy Foundation

Louise Matteoni
Regent
New York State Board of Regents

Paula H. Mayhew
Associate Director
Middle States Association

Herman L. Moore
Director, Regional Government Affairs
Georgia-Pacific Corp.

David A. Nasby
Director, Community Affairs and Vice President
The General Mills Foundation, General Mills, Inc.

Michael T. Nettles
Vice President for Assessment
University of Tennessee, Knoxville

Paul M. Ostergard
Vice President, Director of Corporate Contributions
Citibank

Linda K. Peek
Director, Strategic Communications
The Coca-Cola Company

Irwin H. Polishook
President, Professional Staff Congress
City University of New York

Sheila S. Polishook
Chairman, History Department, Queensborough Community College
City University of New York

Werner Rogers
Georgia Superintendent of Schools
Georgia Department of Education

Alan Rosenberg, Esq.
Chairman
The New York Alliance for the Public Schools

Robert Rothman
Associate Editor
Education Week

Linda Shirley
Project Coordinator
National Dropout Prevention Center

Karl Jan Solstad
Director of Education, School of Education
University of Colorado

Ruth S. Steele
Superintendent of Schools
Little Rock School District, Arkansas

Floraline I. Stevens
Director, Evaluation and Assessment
Los Angeles Unified School District

Herb Sussman
President Emeritus
Rio Hondo Community College, California

Nancy S. Ticktin
Consultant
Bankers Trust Company

George E. Uhlig
Dean, College of Education
University of South Alabama

Sheila W. Valencia
Associate Professor, School of Education
University of Washington

Edmond L. Volpe
President, The College of Staten Island
City University of New York

William "Sonny" Walker
Vice President
National Alliance of Business, Atlanta

Gloria Weiner
Weichert Realtors
Larchmont, New York

Max Weiner
Dean, Graduate School of Education
Fordham University at Lincoln Center

Franklin P. Wilbur
Director of Undergraduate Studies
Syracuse University

Patricia L. Willis
Executive Director
BellSouth Foundation

Arnold Wright
Vice President/Executive Director
CIGNA Corporation/CIGNA Foundation

# THE NATIONAL EDUCATION GOALS REPORT*

Pascal D. Forgione, Jr.
Executive Director
The National Education Goals Panel

Americans believe strongly in the power of education. Almost 90 percent, according to a recent Gallup Poll, rate a world-class education system as critical to the nation's future. More than 80 percent believe that local public schools should conform to national standards of achievement. And yet the evidence today is clear that we do not yet have a world-class system.

A mere 14 percent of American eighth graders can solve problems involving fractions, decimals, percentages, or simple algebra.

A third of eighth graders are unable to figure the cost of a meal from a menu.

A third of eleventh graders cannot write a coherent paragraph about themselves.

If the nation is to measure up to the technical and economic demands of the next century, we must all get involved -- public officials, educators, parents, business and community leaders and students alike -- to meet this challenge. If we want the best education system in the world, we must transform ourselves, young and old, into a nation of learners.

Toward this end, the nation's Governors and the President of the United States have launched an unprecedented ten-year effort. At the heart of this effort are six National Education Goals -- six ambitious targets for us to reach.

*Reprinted from the "Executive Summary," The National Education Goals Report, 1991.

Much as corporate officers regularly report their progress to stockholders, the National Education Goals effort calls for annual reports to be made to the American public and their elected officials showing how the nation and the states are performing on each of the six Goals. The Governors will also issue individual state reports that assess the outcomes and specific reforms within their states.

The point of these reports is not to create an artificial competition that ranks and compares states with one another. Such an approach will not get us where we need to go. Rather, the annual reports will measure progress against past performance and against world-class standards.

Most Americans are aware that scores on existing national tests have declined or stagnated during the past several years. But few associate such declines with the education of their own children or the performance of their own schools. And why should they? They have been lulled into a false sense of security by report cards showing above-average performance by their children. They assume that schools and students elsewhere must be the problem.

What we as a nation have failed to grasp is that simply being above average in the United States or merely improving upon last year's standings does not mean we have the knowledge or skills to prosper in an arena of increased global competition. We must stop settling for mediocrity. Instead, we must reach higher for world-class standards of educational performance.

This first report establishes the baseline for the National Education Goals -- where we stand today in relation to our six Goals. It shows improvements in some important areas. The percentage of students completing

a high school degree or its equivalent, for example, is at an all-time high. And student drug use overall has declined substantially.

But such gains pale in comparison to our educational shortcomings. New data released for this report reveal that fewer than one in five students today can be considered competent in mathematics. American students also rank significantly behind their peers in other advanced countries in science. Even worse, information is simply not available in many cases to tell us where we stand. We lack data on critical topics such as young children's readiness for school, adults' work-force skills, and what college graduates know and can do. Accurate and uniform information on the baseline performance of states was also not available for many of the Goals.

Filling these gaps in our knowledge is as important for forming an action agenda as focusing on those areas where data are available. Future reports must replace these voids with direct measures to tell us whether we are doing what we must to reach our Goals.

Given the distance that still separates us from world-class performance, this report sounds a call to action. It challenges us to translate an abstract belief in the power of education into actions -- policies, practices, and habits that value learning. It challenges us to become a nation of learners.

GOAL 1 - READINESS FOR SCHOOL: By the year 2000, all children in America will start school ready to learn.

All children are born with the desire to learn. But they need to develop in several important ways to succeed in their first year of school. They must be healthy and be able to participate in school activities. They

must possess the social and emotional skills to get along with others in the classroom. They must be able to use language to communicate effectively with other children and adults. And they must have some general knowledge and understanding of the immediate world around them.

The encouragement of parents and a good beginning in school are critical to all future learning. Children who succeed in the early years of school bring a set of attitudes and behaviors -- such as creativity, independence, and curiosity -- that foster further learning. When children falter in their first classroom experiences, however, they can fall behind in class and begin to withdraw from the world of learning. Problems multiply as academic failure, negative behavior, and low expectations feed on one another.

Unfortunately, though we know which qualities promote success in kindergarten and sustain further learning, we have never attempted to assess systematically the degree to which young children in the United States are ready for school. Measuring the social and emotional well-being -- and even general knowledge -- of 3-to 5-year-olds is a considerable challenge.

OBJECTIVES:

- All disadvantaged and disabled children will have access to high quality developmentally appropriate preschool programs that help prepare children for school.
- Every parent in America will be a child's first teacher and devote time each day helping his or her preschool child learn; parents will have access to the training and support they need.
- Children will receive the nutrition and health care needed to arrive at school with healthy minds and bodies, and the number of low-birthweight babies will be significantly reduced through enhanced prenatal health systems.

WHAT WE NOW KNOW

Because we currently lack the tools to measure readiness for school, we must report a grade of incomplete on the first of the nation's six National Education Goals. This is in the face of significant concern that many children are at risk of entering their first year of school without the skills necessary for classroom learning.

WHAT WE STILL NEED TO KNOW

In future years, the National Education Goals Panel is commited to finding accurate, direct measures of children's readiness for school -- in terms of their physical, emotional, social, and intellectual development -- to fill the void in our current base of knowledge. The technical challenges involved in properly assessing young children are many. There are also valid concerns about minimizing the burdens of testing on children and ensuring that results be used to create a national profile and not to place individual children. We believe it is not only possible to create such a measurement system, it is necessary. Having reliable information about young children's readiness for school is essential to ensure that all children have an opportunity to succeed as part of a nation of learners.

GOAL 2 - HIGH SCHOOL COMPLETION: By the year 2000, the high school graduation rate will increase to at least 90 percent.

To become a true nation of learners, we must encourage all of our children to finish high school. Youths who drop out face the prospect of low-paying jobs, chronic unemployment, and limited opportunities for advancement. They can expect to earn $200,000 less during their lifetimes than students who graduate. And these lost earnings can reach into the

billions of each class of students that moves through our schools. Living on the edge of the nation's economy, such youths are susceptible to teen pregnancy, crime, and despair.

Dropouts also inflict substantial burdens on society. Their diminished earnings and restricted opportunities reduce tax revenues and generate added social costs for such purposes as public assistance, medical care, and crime prevention.

In the future, new technologies and job requirements will further increase the skill requirements of workers. Though a high school degree is no guarantee of competence, the lack of one is a clear sign that a person does not have the skills and knowledge needed in the workplace.

The percentage of students completing high school has been inching up in recent years, especially for Black students. However, we need to make much greater gains in the 1990s than we did during the 1980s to achieve our goal of a 90 percent completion rate.

Dropping out is also a reversible decision. Youths can return to a regular high school or complete their degrees by enrolling in alternative programs of study and obtaining high school credentials. The General Educational Development (GED) program, for instance, gives those who have not completed school the opportunity to change the future by earning a high school equivalency certificate.

OBJECTIVES:

-The nation must dramatically reduce its dropout rate, and 75 percent of those students who do drop out will successfully complete a high school degree or its equivalent.

-The gap in high school graduation rates between American students from minority backgrounds and their non-minority counterparts will be eliminated.

WHAT WE NOW KNOW

-Between 1975 and 1990, the number of all 19- and 20-year-old students completing a high school degree or its equivalent rose from 81 to 83 percent.

-Among Blacks, the completion rate rose from 66 to 78 percent during roughly the same period.

-Hispanic completion rates, however, have remained essentially unchanged at approximately 60 percent.

-Nearly half of the 1980 sophomores who dropped out returned and completed high school within the following six years.

WHAT WE STILL NEED TO KNOW

We still need reliable, consistent information about how many youths drop out and fail to complete high school for each state. Current state data are not uniform or comparable. We must remedy inaccuracies and establish common definitions throughout the country of those who complete high school and those who drop out.

GOAL 3 - STUDENT ACHIEVEMENT AND CITIZENSHIP: By the year 2000, American students will leave grades four, eight, and twelve having demonstrated competency in challenging subject matter, including English, mathematics, science, history, and geography; and every school in America will ensure that all students learn to use their minds well, so they may be prepared for responsible citizenship, further learning, and productive employment in our modern economy.

Creating a world-class education system means finding out what students actually know and can do. But determining this is not so simple. The kind of information needed goes beyond traditional methods of testing and reporting. We need to specify our expectations for student performance, making sure that they are high enough to match the highest levels in the

world and we need to determine how many students meet these expectations.

The third National Education Goal focuses on the five subjects of English, science, mathematics, history, and geography. They represent the core content that all students need to master as they move through formal school -- the subjects that provide them with a foundation for future academic study and competence in their subsequent jobs. This report reveals for the first time the gap between how our students actually perform in one of these subjects -- mathematics -- and how they should perform to be considered competent for their grade. This gap is enormous.

While these five subjects constitute an important part of what schools should offer students, they are not all that schools need to offer. A full appreciation of the fine arts and mastery of one or more foreign languages are examples of additional competencies that our schools must foster if we are to produce fully educated and well-rounded citizens of tomorrow.

We are also reporting some disappointing information about signs of good citizenship among young Americans. Citizenship represents another outcome of schooling that is, in many ways, as essential for our future as is competence in subject areas. The social, ethical, and political challenges that confront Americans today, as well as the diversity of our people, make it more important than ever that, as a nation of learners, we become an informed and involved people.

OBJECTIVES:

-The academic performance of elementary and secondary students will increase significantly in every quartile, and the distribution of minority students in each level will more closely reflect the student population as a whole.

-The percentage of students who demonstrate the ability to reason, solve problems, apply knowledge, and write and communicate effectively will increase substantially.

-All students will be involved in activities that promote and demonstrate good citizenship, community service, and personal responsibility.

-The percentage of students who are competent in more than one language will substantially increase.

-All students will be knowledgeable about the diverse cultural heritage of this nation and about the world community.

WHAT WE NOW KNOW

-Fewer than one in five students in grades four, eight, and twelve has reached the competency standard in mathematics.

-However, the number of Advanced Placement exams taken by high school juniors and seniors in English, mathematics, science, and history has increased by 51 percent since 1986.

-Only 48 percent of young adults between 18 and 20 years of age are registered to vote, compared to 70 percent of all U.S. Citizens 18 or older.

WHAT WE STILL NEED TO KNOW

We have only a limited profile so far of student mastery of the five subjects in our third Goal. We know the most about proficiency in mathematics, where an initial standard of competence has been established. But no such standards of performance exist in the other subject areas. We need to establish these standards, along with a voluntary national system of examinations (not a single national test) to tell us whether or not the standards are being met. Only then will we see where our educational efforts are succeeding and where they are falling short.

GOAL 4 - SCIENCE AND MATHEMATICS: By the year 2000, U.S. students will be first in the world in science and mathematics achievement.

As the world becomes increasingly competitive, it is essential for all

Americans to be proficient in science and mathematics. Mastery of these subjects stimulates innovation across all sectors of our society and spawns discoveries that aid in our nation's security and benefit our environment. For these reasons we have set a separate Goal that our students will be second to none in these areas.

Unfortunately, data from international assessments conducted during the 1980s point to relatively poor achievement by American students compared with youths in other countries.

Achievement in science and mathematics is not limited to those who are born with special talents. In a nation of learners every citizen should demonstrate competence in these subjects. Indeed, our quality of life is endangered when citizens cannot make informed decisions on issues that are increasingly shaped by science, mathematics, and related technologies. Fluency in science and mathematics gives an individual powerful tools with which to adapt in a world of rapidly changing knowledge.

OBJECTIVES:

- -Math and science education will be strengthened throughout the system, especially in the early grades.
- -The number of teachers with a substantive background in mathematics and science will increase by 50 percent.
- -The number of U.S. undergraduates and graduate students, especially women and minorities, who complete degrees in mathematics, science, and engineering will increase significantly.

WHAT WE NOW KNOW

- -In a five-nation study of 13-year-olds in 1988, U.S. students ranked fourth in science achievement and last in mathematics.
- -American 14-year-olds fell near the bottom of a 17-nation study of science achievement conducted between 1983 and 1986.

-Students from 12 nations significantly out performed U.S. 13-year-olds in one or more areas of mathematics in a 1980-1982 study.

WHAT WE STILL NEED TO KNOW

We need more current data on international achievement in science and mathematics. Results from the first of several new surveys to be conducted between now and the year 2000 will be reported next year. We also need better information about methods used to teach science and mathematics in this country and how they differ from methods used in nations where performance is better than ours.

GOAL 5 - ADULT LITERACY AND LIFELONG LEARNING: By the year 2000, every adult American will be literate and will possess the knowledge and skills necessary to compete in a global economy and exercise the rights and responsibilities of citizenship.

As we become a nation of learners, we must pay as much attention to the education of adults as we do to that of children. Americans tend to associate adult education with acquiring minimum literacy skills such as reading a newspaper, interpreting a map, or performing simple arithmetic. And it is true that we must ensure that all Americans are functionally literate. But this target alone is not good enough if we are to prosper in the 21st century. We must all add to our skills.

Nearly 85 percent of America's workers for the year 2000 are already in the work force. The knowledge and skills they possess will be crucial factors in our success or failure in the global marketplace. Tomorrow's standard of living thus depends on the education and training of today's workers.

Recent surveys, however, suggest that while most Americans have the most basic fundamental literacy skills, far fewer are able to perform more

complex tasks that require them to process and synthesize many pieces of information.

Our future also depends on the performance of America's higher education system. Our colleges and universities are this nation's prime knowledge engine. They need to produce graduates who are as learned as those anywhere in the world.

OBJECTIVES:

-Every major American business will be involved in strengthening the connection between education and work.

-All workers will have the opportunity to acquire the knowledge and skills, from basic to highly technical, needed to adapt to emerging new technologies, work methods, and markets through public and private educational, vocational, technical, workplace, or other programs.

-The number of quality programs, including those at libraries, that are designed to serve more effectively the needs of the growing number of part time and mid-career students will increase substantially.

-The proportion of those qualified students (especially minorities) who enter college, who complete at least two years, and who complete their degree programs will increase substantially.

-The proportion of college graduates who demonstrate an advanced ability to think critically communicate effectively, and solve problems will increase substantially.

WHAT WE NOW KNOW

-In two surveys of American adults, the vast majority were able to perform tasks such as locating a particular fact from a newspaper article, matching a money-coupon to a shopping list, or adding two entries on a bank deposit slip.

-But most adults were unable to carry out more complex tasks such as interpreting the main argument from a lengthy newspaper column, using a bus schedule to determine how long it would be before the next bus arrived at a desired destination, or using a page from a shopping catalog to fill out an order form and then calculate the costs of a number of items.

WHAT WE STILL NEED TO KNOW

We are still largely ignorant about the nation's standing relative to

this Goal. We need more recent data on the literacy levels of all adults for the nation and for individual states. (Such information is expected in 1993.) We also need state and national data on how well the knowledge, skills, and attitudes of American workers compare to those of workers in other nations, especially those that are our prime economic competitors. And we need to know what our college students have learned by the time they graduate.

GOAL 6 - SAFE, DISCIPLINED, AND DRUG-FREE SCHOOLS: By the year 2000, every school in America will be free of drugs and violence and will offer a disciplined environment conducive to learning.

For our schools to fulfill their vital educational mission, they must provide teachers and students with an environment that is safe, disciplined, and drug-free. No school can function effectively in an atmosphere where illicit drugs are commonly sold or used, or where crimes against teachers and students are pervasive, or where classes are regularly disrupted.

Recent data show that the nation has made some progress in this area during the 1980s. But there is still much to be done.

Creating safe, disciplined, and drug-free schools will not guarantee that more students complete high school, or that academic achievement will increase. But we have little hope of reaching these other Goals if we do not achieve Goal 6.

OBJECTIVES:

-Every school will implement a firm and fair policy on use, possession, and distribution of drugs and alcohol.

-Parents, businesses, and community organizations will work together to ensure that schools are a safe haven for all children.

-Every school district will develop a comprehensive K-12 drug and alcohol prevention education program. Drug and alcohol curriculum should be taught as an integral part of health education. In addition, community-based teams should be organized to provide students and teachers with needed support.

WHAT WE NOW KNOW

-Marijuana use at school by 12th graders dropped from 21 percent in 1980 to 6 percent in 1990.

-Nevertheless, in 1989, about three in ten 12th graders reported that alcohol and marijuana were easy to obtain in their schools.

-In a recent survey of crime and violent acts, 42 percent of high school seniors reported having had property stolen at school, 14 percent reported being injured without a weapon being used, and 6 percent said they were injured with a weapon.

-Black students are much more likely than Whites or Hispanics to be victims of violent acts at school involving weapons.

-The vast majority of public school teachers feel safe in their schools, though in a recent survey eight percent reported being threatened with injury and two percent reported being physically attacked during a 12-month period.

-Teachers in cities are more likely than those in other communities to feel unsafe in their schools (especially after school hours) and to report being verbally abused or threatened with injury.

WHAT WE STILL NEED TO KNOW

We need better information at the state level about how safe, disciplined, and drug-free schools are, and we need these data to be collected and reported regularly and in a comparable manner by each state.

# ASSESSING THE BUSINESS ROUNDTABLE'S 1990s EDUCATION INITIATIVES

Michael T. Nettles
Vice President for Assessment
University of Tennessee

The Nation's Business Roundtable (BRT) has produced a plan to contribute significantly to national efforts at making systemic changes to improve America's schools. Each of the approximately 175 CEOs who comprise the BRT is working with one governor from a state that he/she selects to set goals and implement a specific agenda of programs designed to improve and strengthen education in that state. Jointly, the governor and CEO along with others are setting goals and planning strategies to affect the course of education in a state over the next decade. The BRT's plan is called "The BRT Educational Agenda: Essential Components of a Successful Education System."

The nine essential components of the BRT plan are the following:

1. The new system will be committed to four operating assumptions:

   -All students can learn at significantly higher levels.
   -We know how to teach all students successfully.
   -Curriculum content must reflect high expectations for all students, but instructional time and strategies may vary to assure success.

2. The new system will be performance or outcome based.

3. Assessment strategies must be as strong and rich as the expected outcomes.

4. Schools should receive rewards for success, assistance to improve, and penalties for failure.

5. School-based staff will have a major role in making instructional decisions.

6. Major emphasis will be placed on staff development.

7. A high-quality pre-kindergarten program will be established, at least for all disadvantaged students.

8. Health and other social services must be sufficient to reduce significant barriers to learning.

9. Technology will be used to raise student and teacher productivity and to expand access to learning.

The University of Tennessee has assisted the BRT by helping to evaluate the effectiveness of its educational initiatives. The first annual status report has two components. The first component assesses the efforts by the CEOs in the following five categories:

1. Coalition Formed/Joined: Was a coalition of business leaders formed within the state whose main purpose was to pursue education reform? Or, alternatively, did business leaders join an existing coalition with a similar purpose?

2. CEO/Governor Meeting: Did the current Governor and one or more of The Business Roundtable's CEOs or designees meet to discuss education reform?

3. Coalition Established Comprehensive Agenda: Did the coalition establish a reform agenda related to educational concerns?

4. Comprehensive Agenda Including Business Roundtable Essential Components: Did the coalition commit to pursue systemic restructuring by adopting The Business Roundtable's Essential Components?

5. Coalition Developed Strategic Plan: Did the coalition develop a strategic plan for implementing The Business Roundtable's Essential Components?

This assessment is presented in a matrix fashion, ranking and comparing the 50 states based upon the percent of the activities completed during the first year. These actions were judged by the CEOs to be an important first step for them to enter the education reform movements in the states; therefore, they decided to measure their activities in these five important categories during the first year rather than prematurely monitoring

the extent to which the states have adopted and implemented their entire agenda.

The data for the matrix were gathered by The University of Tennessee and BRT staffs working together to document events and activities of the CEOs. Each CEO has an appointed staff designee from her/his corporation who may devote as much as full-time to this education initiative. These staff designees keep detailed documentation on the events and activities of their state.

The second component of the 1991 status report is a paragraph about each state's education reform initiatives and the business communities' involvement in the overall state reforms. This component reflects the CEOs' interest in and sensitivity to the larger efforts in the states. Such efforts might include all or part of the BRT agenda, but would not be made up exclusively of its initiativies. In general these paragraphs reveal several trends, but especially the following five:

1. the six National Education Goals that were established by the President and the Governors are being endorsed;

2. emphasis in most states is being placed on results and accountability;

3. budget constraints have slowed reform efforts in many states;

4. in many states, the reform efforts are being led by the Governors and legislative officials, who are seeking involvement and support from business groups in the states; in other states, however, the business community has taken the primary leadership role; and,

5. business involvement in reform efforts is varied, with substantial contributions being made in several states.

Future status reports could be expanded to include new actions that the CEOs undertake to introduce and implement their agendas. These actions will

have to be defined. Future status reports could also monitor the rate at which the states adopt the BRT nine components by establishing policies and laws and could assess as well the relationship that these nine components have to improvements in schools and improvements in student learning and development. These latter types of assessments are ultimately the most important. The data and evidence needed for these assessments may also need to be focused and may even need to be gathered from case studies. Otherwise, given the many education reforms underway in every state, it could be impossible using ordinary large-scale assessment techniques to isolate the unique contribution that the BRT initiatives make to overall change and improvement.

# NATIONAL ASSESSMENT FOR EDUCATIONAL PROGRESS

Archie Lapointe
Executive Director
Educational Testing Service, Princeton

The recent accumulation of statistics has fueled a debate that grows increasingly impatient and frustrated. There are those who argue against more "education bashing," and they are weary of yet another document that relates the same bad news. There are others who are turning over the rocks as they fall and, with a bit more creativity, are discovering that some of the bad news is susceptible to challenge. Finally, there are still others who refuse to accept the seeming inevitability of the trends and who are determined to do something to reverse them. These are often people from outside the education establishment.

Let us consider where all these data come from and the various forms of assessments involved.

First there are the local tests. In addition to teacher-made quizzes and end-of-unit tests administered to support instruction, there are standardized achievement tests purchased from publishers or created by the state departments of education. In a few cases there are penalties attached to the results from these tests in that students who fail to meet certain passing grades are asked to take summer remedial programs, are not permitted to be promoted to the next grade, and in some states, are not graduated from high school. In the vast majority of cases, however, the data, literally tons of it, are accumulated in files, discussed for a few seconds during parent-teacher meetings, and occasionally studied by guidance counselors while they

advise students or talk to parents about college selection or remedial courses.

It has become common practice for these state-mandated tests to be used to rank schools and school districts. The purpose of this exercise is to stimulate a competitive spirit and to encourage principals and teachers to try harder to be number one. I know of no instance where this practice has worked successfully.

In addition to these more informal tests, there are, of course, the college entrance examinations, primarily the American College Testing Program (ACT) and the Scholastic Aptitude Test (SAT), which demand more serious attention by the two million or so high school juniors or seniors who take them each year, and whose results often determine students' success in gaining acceptance at the college or university of their choice.

Finally, there is the National Assessment of Educational Progress (NAEP), which measures a national sample of students at grades four, eight, and twelve and identifies the "body temperature" of American education to see how well students are learning. In 1990, states that volunteered to do so, could administer the National Assessment in mathematics to a representative sample of their eighth grade students and compare their performance to that of other states. In 1992, 43 or 44 states will do the same thing for reading and mathematics at the fourth and eighth grade levels. This testing is an experiment authorized by Congress, and, after considering the results of the trial, the NAEP legislation will be reexamined regarding whether or not to extend this practice.

Facing all of this testing, teachers, principals, and superintendents are beginning to say "enough!" They claim it is disruptive. And they are

right. They assert it takes too much time from teaching and learning. And they are right. They claim that the results are not being used effectively. And they are right. The protests are growing louder and more shrill as the nation's answer to most of education's problems these days seems to be to create and administer yet another test. The "America 2000" plan to improve education has as a key feature the creation of national achievement tests to be administered to every student at three grade levels in five key subject areas. The intention is to provide scores for individual children, results at the school level and at the district level, as well as information that will permit state departments of education to monitor their own progress and contribute to national statistics. The President and the Secretary of Education think this testing is a good idea. The Congress disagrees.

How did we get to where we are? If one examines closely the origins of this particular Hell we have created, one finds nothing but good intentions. Teacher-made quizzes and tests probably have the noblest objectives of all: to find out if individual students have grasped material and to determine if students can perform certain higher-level skill operations. The use of standardized achievement tests by these same teachers in these same school districts is an attempt to provide teachers with higher-quality test items, with more valid and reliable test instruments, and with diagnostic information for each individual child that will confirm the teachers' own findings or challenge their prejudices about individual children. Most of the states have also installed minimum competence testing programs in order to insure that all children in the state, regardless of their gender, ethnic background, or economic condition are being provided an equal opportunity for a decent education. These are clearly good objectives.

The college entrance examinations were created about 40 years ago in order to provide a more equitable way of evaluating the secondary school performance of college and university applicants. The variation in the quality of American education made reliance on school scores and grades a confusing problem. It is obvious to all who have attended these institutions that an "A" from one secondary school may be equivalent to a "B" from another. The college entrance tests, administered in standardized ways to all applicants, provide another way of comparing college applicants. This effort to be fair and equitable is certainly to be commended.

Finally, the National Assessment was created some 22 years ago by the then Commissioner of Education, Frank Keppel, who was attempting to respond to his congressional mandate to report periodically on the "quality of American education." He asked Ralph Tyler and John Tookey to devise a nonthreatening, unobtrusive set of instruments that would report anonymously on "What American 9-, 13-, and 17-year-olds know and what they can do."

This brief review, which leaves out many customized and specialized testing instruments, is enough to guarantee that there are data available to measure just about any facet of American education. Each test yields a report. Many of them are summarized in periodic cards to parents. These results are given attention in local newspapers, and the results of the college entrance examinations and of the National Assessments are increasingly featured in the nation's national media and press.

Some of the results are more valid and reliable than others. Much of the data, like most statistics, is misused or misinterpreted, and very little of it is examined to discover the positive trends that are often either very near the surface or clearly exposed.

In summary, there are many forms of assessment designed to serve many purposes, each form producing more data than we know how to digest, and the country's plan seems to be to add more testing to the educational mix. But apart from this proliferation of testing, what is the main issue? Quite simply and emphatically the need for standards.

If one were to ask the question,"Do we need local, state, or national standards?" the answer would be, "Yes." I can easily describe another list of issues facing American education, such as a national test, a national curriculum, school financing, integration, school choice, restructuring American schools, and the qualifications of teachers. In the absence of generally accepted standards, i.e., of some clear idea of where we want to go or of what we want the schools to do, it is very difficult to even begin evaluating these other issues.

Over the years, American schools have been successful. They are probably among the most responsive institutions in our society. Forty years ago, colleges and universities asked them to prepare students by teaching Latin and Greek. They did. Thirty years ago, after Russia sent up Sputnik, schools were asked to teach more mathematics, more science, and more foreign languages. They did. Twenty years ago, schools were instructed to have children implement the laws on integration with which we adults had failed to comply. They did. During the last fifteen years, schools have been encouraged to offer what some have called "cafeteria-type" education programs to accommodate needs, interests, and, alas, lack of motivation of many students. They have. In a few months, some districts in America will be asked to pass out contraceptive devices to their high school students. They will.

There is an unbroken record here of schools doing what the public tells them to do. Sometimes we have told them directly with laws and regulations. Other times it has been the lack of clear instruction that has influenced their activity. But schools cannot be faulted for not living up to what is expected of them.

One example, among others, will point to successes that have been little acclaimed or unnoticed within the general furor that has characterized the education debate. During the past twenty years, the gap separating the performance of Black, Hispanic and White students has diminished dramatically at all three age levels measured by NAEP. This development suggests that twenty years ago when we told the schools to focus on the disadvantaged and minority populations and we told them to pay attention to "the basics," not only did they do it, but they were successful in ways that are clearly measurable.

Among all the debates and discussions going on at the present time, in Washington, in state capitals, at the school-board level, and within professional associations, I think those that involve a serious consideration of a setting of standards will be the most fruitful. I would strongly recommend that until we have a clear idea of what those standards should be, an immense task in and of itself, we lay aside all the other discussions about how we might reach our goals and about how we might measure progress along the way.

The original education objectives in each of the American states that set up common schools were generally clear and simple. The natural confusion that results from 300 years of a glorious, multi-cultural history is that we can no longer recognize what the schools are expected to do.

School board members themselves are usually unable to present a statement of their educational objectives that goes beyond the kind of meaningless rhetoric that cannot be translated into specific policies or actions.

After considering these sobering ideas, what can one recommend? I said earlier that I believe American schools have been successful in the past. I also agree with the severest critics that our schools have to get much better very fast, and I would recommend a three-part strategy. All of these elements are in place right now, some of them in encouraging stages of development.

First, let us establish some clear standards. This is a very tough assignment. In the fullest meaning of the expression, it has to involve a national dialogue. President Bush and the Governors have made a truly remarkable start. In many of the states, business and school leaders have established very strong organizations for implementing the National Goals. But while a National Goal can read simply "Becoming number one in mathematics and science," the fourth-grade teacher needs something more specific to guide his/her instruction. The world is always a little surprised that fourth grade teachers want to play a role in determining what teachers should expect from their students. If we want instructional personnel to implement any strategies for moving toward our goals, teachers and school principals must be heavily involved in debating and establishing just what those goals mean and what will be required in the three million classrooms across the country.

Once standards are in place, discussions involving how we achieve them will fall more clearly into perspective. Each debate will now be framed within a context against which conclusions can be evaluated. The questions

involving children arriving in first grade in good health will answer themselves. The unacceptability of drug-infested school buildings in an environment attempting to achieve clear goals will be self-evident. And, given the unequivocal record established by schools in responding to clear calls for action, it can be assumed that local teachers and administrators will find ways to deliver what is expected of them.

Once the goals are set in place, clearly understood and accepted, there needs to be a simple, nonduplicative set of assessment techniques for measuring progress toward those goals. Today it is not uncommon for an individual fourth grader to be asked a series of questions in mathematics by a teacher to see what the youngster knows and can do; to be asked a similar series of mathematics questions during a standardized test to provide information to his school district; to be asked a third set of questions on a state test to provide state-level policymakers with the same kind of information. That same student may also be part of a sample that will provide data to NAEP on what American Fourth graders know and can do in mathematics. If objectives are clear and compatible, it seems reasonable to expect that a series of evaluation techniques can be combined so that results can be provided to all those who need them.

I have deliberately moved away from the use of the words "tests and examinations" and am referring to the notion of an assessment or evaluation system. A key part of the system has to be the teacher who spends several hundred hours with a student each year. As a trained professional, he/she should be expected to form judgments about individual strengths and weaknesses.

These impressions should be formalized and probably illustrated with a

portfolio of student work when they are reported to principals and parents. They should also be confirmed by some kind of outside evaluation exercise, probably in the form of a test, an examination, or a series of performance tasks that are used in a broader environment, for example, across the state or across the country, as in NAEP, so that the individual teacher and the individual student will have some understanding of how their work compares with that of other students in other states and even with students in other countries.

My last recommendation has to do with reporting results. Certain principles provide a foundation for any consideration of disseminating information about educational testing:

- -Any test or assessment device used should be unbiased and of the highest quality possible.
- -Any data analysis of the results should be based on solid, accepted research.
- -Any results identified should be honest and unbiased.
- -All results should be analyzed by experts with a wide range of opinions about their significance.

The major problems we encounter in any consideration of testing have little to do with these principles. The industry, the practitioners, the media are all fascinated by the process of testing. There is an enormous amount of detail involved in the construction of test questions, in the administration of the instruments themselves, in the analysis of the data, and in the development of reports. My personal view is that the whole exercise should be viewed as nothing more than an information system designed uniquely and specifically to provide information, not data, to practitioners, parents, and policymakers.

The difficulty has to do with the clarity and understandability of testing reports. Admittedly, it is difficult to present this information in clear and understandable ways because inevitably findings are based on estimates drawn from examples or samples of performance. A hundred-question mathematics test, for instance, does not measure everything there is to know about mathematics. It is only a sample of those things. Therefore, when someone reports that a student received a 70 percent score on a hundred-question math test, one can only estimate what an individual truly knows about all of mathematics.

This difficulty is why information has to be reported carefully, but reporting can still be done clearly, understandably, and without a lot of footnotes and unnecessary caveats. Our lives are full of judgments that are based on partial data, and, by and large, with ordinary prudence, we are able to manage and be successful. Using educational data is no different.

In summary, if we are successful in achieving some of these objectives, it may be that our greeting the assessment messenger in the future will reflect a more enthusiastic attitude -- a bit like checking the scores on a sports page to see how our favorite team is progressing toward the pennant. We could have a great season. All of the fans clearly want us to win. The players have already demonstrated they can. What we need is solid management and consistent front-office support. The combined forces at work to make education better will be served well by the progress in national educational assessment now under way.

# GEORGIA STATE ASSESSMENT PROGRAM*

Stanley Bernknopf
Division of Assessment
Georgia Department of Education

The Georgia State Board of Education is committed to gauging the quality of schooling occurring in Georgia public schools. At the same time, the Board is committed to a restructuring of education in order to better meet the needs of students and communities in the Georgia of the 1990s and into the twenty-first century. The state's accountability requirements must enhance and not impede change.

Assessment instruments do not truly serve a monitoring and accountability function unless they measure what the schools are actually supposed to teach. The State Board believes that the accountability system in Georgia schools should be based primarily on student achievement in the content defined by the state's Quality Core Curriculum (QCC). Assessment must match instruction in level, in scope, and in methodology. Instructional priorities should determine what is assessed. If, for example, higher-level thinking skills are an instructional goal, then students must be asked to demonstrate such skills. The assessment strategies chosen to measure progress toward a goal will clearly influence how students are taught. This implies that goals for student learning must be clearly defined and that assessment strategies must be as varied and as rich as the educational outcomes desired.

*This paper is a summary of a larger position paper prepared by Georgia Department of Education.

Local school districts must assume responsibility for formative evaluations of student learning; the state, nevertheless, should remain responsible for summative evaluations that take place at certain key points in schooling, such as at the end of primary, elementary and middle grades. The state also should have primary responsibility for certifying the status of enrolled high school students with respect to certain student outcomes.

Providing outcome information for program evaluation through a testing program, however, does not constitute an accountability program. Accountability, by definition, implies responsibility. An educational accountability structure must include responsibility for goals established to reform the learning of students. The state's assessment program, therefore, must provide direction for change, and the state must ensure that reform occurs by providing resources, technical assistance, and leadership to systems and schools whose educational achievement is below standard. Those responsible for student learning -- schools and systems -- must plan and implement reform strategies to ensure that learning occurs and that students make progress toward instructional goals.

The state's assessment program should provide three kinds of accountability information:

1. the standing of schools and systems with respect to state instructional goals and the curriculum determined to be relevant to those goals;

2. the achievement of schools and systems with respect to national student performance;

3. the success of individual students with respect to established goals at the time of high school graduation.

To accomplish this aim, Georgia's assessment program has moved away from tests that emphasize discrete, fairly low-level skills, such as

Criterion Reference Tests (CRTs); and likewise we must not rely on Norm Reference Tests (NRIs), which may cover a wide range of facts but in a fragmented way that cannot yield a comprehensive view of student achievement.

Assessment must influence instruction toward higher-level skills by "moving from normative tests to substantive criteria ... from disconnected assessments to integration of assessment and teaching." (Council of Chief State School Officers, 1991.)

Restructuring the state's assessment program to focus on schools and systems rather than on individual students will facilitate the measurement of higher-level goals and broader content and skill coverage. Although it is not feasible for a state program to evaluate every student on every objective, schools and systems can be evaluated on the extent to which they are implementing the entire range of QCC objectives rather than solely on student achievement of basic reading and mathematics skills.

Not every student needs to respond to every objective in order to assess school and system status and provide information to facilitate change. The appropriate sampling technique, i.e., matrix sampling, of student responses to different test forms will allow for the increased content coverage and, at the same time, require minimal testing time for individual students. Results across test forms can be aggregated at the school and district level to obtain a valid picture of student performance. The matrix sampling approach also serves to protect the integrity of test results because individual students will take one of a number of test forms made up of different questions. Students can be "prepared for testing" only through being taught the entire curriculum.

The state assessment program displayed in the appendix is compatible with the goals and purposes of a restructured educational system. The program meets a variety of needs for information, minimizes state-required testing time, and allows flexibility to local education units in meeting local needs and priorities for assessment. The proposed program provides for individual student assessment of readiness at the beginning of schooling and for certification that students have acquired a set of valued skills and concepts prior to high school graduation.

Local districts and individual schools will be monitored by assessing cohorts of students at key points: end of primary grades, end of elementary grades, end of middle grades and near the end of high school. This assessment will utilize a series of measures based on Georgia's required curriculum. The QCC is not only a description of content; it offers a perspective on a variety of ways of understanding the content. The testing instruments will be a combination of commercially available items in which students are required to demonstrate that they can integrate and apply certain skills and knowledge and more traditional test items covering a broad range of content.

The Georgia testing structure will assess higher-order thinking skills as well as basic skills and concepts. School and system performance on nationally norm-referenced measures also will be assessed. Both the curriculum-based items and norm-referenced items will be combined in a series of test forms administered to samples of students at particular grade levels in each school. Information will be collected and provided in a way that leads to corrective action or positive change to enhance student progress.

The State of Georgia is committed to quality standards for every student

enrolled in the public schools. In addition to those factors mentioned above, this commitment calls for substantial changes in the preparation of teachers, the way instruction takes place at each school site, and a system of measurement that evaluates what students are really expected to master. Our assessments must be geared to reveal what is occurring in the educational process. Going into the next century, our expectation is to inculcate educational success that is linked directly to national and world standards.

## APPENDIX

### PROPOSED STATE ASSESSMENT PROGRAM

K READINESS ASSESSMENT (GKAP)
- -Developmentally appropriate, instruction-based
- -All public school kindergarten students

3 END OF PRIMARY GRADES ASSESSMENT
- -Curriculum-based assessments
- -Norm-referenced assessments
- -Administered through matrix sampling
- -Individually administered writing assessment

Approximate time - 3 hours

5 END OF ELEMENTARY GRADES ASSESSMENT
- -Curriculum-based assessments
- -Norm-referenced assessments
- -Administered through matrix sampling
- -Individually adminstered writing assessment

Approximate time - 4 hours

8 END OF MIDDLE GRADES ASSESSMENT
- -Curriculum-based assessments
- -Norm-referenced through matrix sampling
- -Individually administered writing assessment

Approximate time - 5 hours

11 GRADUATION ASSESSMENT
- -Individual student assessment which will include process and application skills and an extended writing sample
- -May include a matrix sampling of items covering additional QCC content and NRT for accountability

Approximate time - 6 hours

# LOS ANGELES CITY SCHOOL ASSESSMENT PROGRAM

Floraline I. Stevens
Director, Program
Evaluation and Assessment Branch

The Los Angeles Unified School District, with over 650,000 students and increasing by approximately 15,000 students each year, is the second largest public school system in the United States. The district is composed of 648 schools and 187 centers served by 30,428 teachers. According to data from the California Assessment Program (CAP): (1) the number of poor students in the district is increasing; (2) the level of parents' education is decreasing; and (3) the number of limited-English-proficient (LEP) students is increasing. The district has students who speak over 80 languages; however, more than 90 percent of the LEP students speak Spanish. Currently, more than two-thirds of the district's students are Hispanic, 15% are Black, 14% are White, 4% are Asian, and less than 1% are American Indian.

Approximately 142,000 students in 1990-91 were in year-round schools in 77 multi-track and 7 single-track elementary schools, 9 multi-track junior high schools, and 4 multi-track senior high schools. In addition, 578 students in 4 continuation high schools were on year-round schedules. Eighty-seven percent of the year-round schools' student population was Hispanic. For 1991-92, the entire district is on a year-round school calendar with most of the schools having 3 tracks in session for 90 days and on vacation for 30 days (a 90/30 schedule). Students on each track are provided with 180 days of instruction.

There are approximatley 269,000 students served in 274 regular Chapter 1 schools. There are 215 elementary schools and 47 secondary schools in the Chapter 1 Program. Not all students in Chapter 1 schools are Chapter 1-eligible. Eligibility is based on poverty level (Aid to Dependent Children and Free-Lunch Program) and educational disadvantagement (at or below the 49th percentile on a norm-referenced test in reading or mathematics). Limited-English-speaking is not a criterion for Chapter 1 eligibility.

The LAUSD magnet program is one of the largest voluntary desegregation programs in the United States. There are 28,000 students enrolled in 95 programs. Evaluation results over the years show that parents and the community believe that students have strong educational opportunities when attending magnet schools. There are, however, 18,000 students on the waiting list each year to attend magnet schools.

Cutting across this array of special programs to meet the needs of the district's students are the issues of student academic achievement, student and teacher accountability, and improvement of instruction. The district's assessment program attempts to address these issues.

Norm-Referenced Testing

The Comprehensive Tests of Basic Skills, Form U (CTBS/U) are norm-referenced tests administered each spring to all students in grades 1-11. Subject areas tested are reading, mathematics, and language. (Kindergarten testing is optional.) Aprenda is a Spanish language norm-referenced test which is administered to all Spanish-speaking students in grades 1-11 who have not met specific criteria related to English reading instruction. The

same subject areas are tested. In addition, for Chapter 1 eligibility-assesment purpose those students who are LEP and whose native languages are Philipino/Tagalog, Armenian, Korean, Vietnamese, and Cantonese/Mandarin are given a truncated norm-referenced test called the BEST (Basic Elementary Seconday Test) in their respective languages.

Norm-referenced testing provides teachers, principals, and state and federal agencies with percentile scores. These scores are used to determine students' academic achievement compared to a national sample. Aggregated, these scores provide comparisons for grade levels, schools and school districts. Percentile scores reveal if the scores are below, at, or above the national averages. These scores do not inform teachers or school administrators where specifically to focus instruction. Criterion-referenced testing fulfills that function.

Criterion-referenced and Competency Testing

Using the CTBS/U as a testing instrument, new psychometrics coupled with technology allow the clustering of items by skill or curriculum objective. This clustering enables items to be rescored for mastery or non-mastery of the skill/objective. For instructional purposes in LAUSD, teachers receive an objectives-based report that shows for each student in the classroom which skills were mastered, approach mastery, or were not mastered.

The same technology allows, with input and selection of specific items by district curriculum specialists, the clustering of items for competency purposes. These items were selected from all subject areas of the CTBS/U that match the designated competency skills for grades 4, 7, and 9.

Secondary students are administered the complete battery of the CTBS/U with the exception of the science test.

Thus, with the administration of one test, three different types of scores are generated: percentiles, percent correct, and pass/fail. High School students who do not pass the competency test in grade 9 have several opportunities to take the test again in grades 10 and 11. For those students who are in grade 12, alternative district-developed tests are administered three to four times in the Winter and Spring. SHARP (reading), TOPICS (mathematics), and WRITING: Sr. (English objective test and writing performance) are the tests. These tests are published commercially by CTB MacMillan/McGraw-Hill.

State Mandated Assessment

Since the 1970s, the state's academic achievement accountability system has been the California Assessment Program. All students in grades 3, 6, 8, and 12 are given a matrix test. That is, a portion of the items covering reading, mathematics, and language is given to each student. Writing performance is measured at grades 8 and 12. At the grade 8 level the assessment has been lengthened to include social studies and science. The CAP was not administered during 1990-91 because of the budget crisis in California, but there are plans for a revised CAP to be administered to grades 4, 5, 8, and 10 with more performance-based items included for 1991-92.

The CAP scores receive much media attention. Each school's scores are printed in The Los Angeles Times, The Daily News, and The San Pedro (Harbor Area) Pilot. No individual student data scores exists for CAP.

Grade level scores are reported for the schools and the district. Therefore, it is difficult to use this information for improvement of academic achieve-ment for specific students, classrooms, or for information to individual teachers.

The revised CAP is expected to introduce individual student scores in some subject areas and grades.

National Assessment for Educational Progress (NAEP)

Specified students in grades 4, 8, and 11 or 12 are administered the NAEP in LAUSD. Because the number of schools and the number of students are small, NAEP is unobtrusive to the district. Furthermore, the administration of the NAEP is handled principally by persons external to the assessment unit and teachers in the schools. Because the NAEP sampling does not have enough students for district scores, the impact of the information is not immediate or substantive in regard to student and teacher accountability for academic achievement.

Student and Teacher Accountability

Academically, norm-referenced test (NRT) scores show that the students in LAUSD are scoring below average in most grade levels in reading, while the mathematics scores are in the average range. A closer review of disaggregated test data taken from integration program evaluation reports shows that there are differences in the academic performances of racial/ethnic groups. African American and Hispanic students score the lowest. Students of all ethnic groups who are in the Magnet School Program show the most academic success, that is, highest percentile scores. The

magnet schools' percentile scores are above the district average and above the national median. Unfortunately, as noted earlier, this program is expensive and has a large waiting list of students.

There is a pilot assessment program that impacts only ten schools in the district. The program is funded with district, state integration, and Chapter 1 monies. It is called the Ten Schools Program. The Ten Schools Program is composed of the ten lowest scoring elementary schools (according to norm-referenced test scores) which had 60 percent or more African American students. The program utilizes effective schools research strategies. In addition, the academic year at these schools has been extended into the summer. Enrichment classes for students have been provided on Saturdays, and after-school tutoring has been available. In an attempt to determine what academic progress students are making throughout the year, a series of interval or segmental tests was developed after teachers, instructional coordinators, and principals met with the district's assessment unit. The school personnel identified those skills they felt were priorities for each grade level. From this information a series of Milepost Tests (named by the teachers) were developed. Three tests per grade level were administered to the students. The CTB MacMillan/McGraw-Hill Company developed and produced the tests with editorial input and revisions coming from the district's Ten Schools Program staff and the assessment unit.

The Milepost testing was originally planned for student accountability and to focus on the skills students needed to learn. However, in addition to these objectives, the test data proved to be valuable for teacher accountability issues. The tests assessed students'

opportunity to learn and the quality of instruction provided by the teachers covering the entire grade level curriculum.

The initial findings from this pilot assessment, along with classroom observations, were that student test percentiles, mastery, and pass/fail scores, are strongly dependent upon opportunity-to-learn issues as well as students' efforts to learn the subject matter.

The district is seriously considering the recommendation of the Office of Instruction and the Program Evaluation and Assessment Branch that interval assessment should be a districtwide program for all students. CAP is also looking at interval assessments.)

For those who wish to support urban public education, there are two fundamental questions: (1) does the instructional program raise academic achievement, and (2) is the instructional program replicable in various academic settings?

Why are those questions important? Our overriding concern is the fact that most urban public school children are coming to school less prepared for learning and have less family support for schooling. In terms of limited financial and human resources, we no longer have the luxury to fund academic programs that do not produce successful outcomes. Most importantly, we want all of our students academically prepared for successful adulthood.

Assessment in its various forms provides many answers relating to these problems. It allows the district to evaluate the learning deficiencies of students throughout their educational progress, and it offers the information needed to decide on what works best in the Los Angeles public schools. We are aware of the constant need for upgrading our assessment

program with regard to performance-based testing and for increasing the use of interval and segmented testing. The Los Angeles Unified School District is committed to a complete assessment system that provides a foundation for teaching and accountability to the public.

# PHILADELPHIA SCHOOL DISTRICT ASSESSMENT PROGRAM

Rita C. Altman
Associate Superintendent
School District of Philadelphia

The educational assessment program in the School District of Philadelphia is very much a "work in progress." I characterize our efforts in this fashion because we have both a retrospective and a prospective history. Our retrospective history is like that of many, if not all, of the urban school districts of our nation. Our prospective history is one of development and discovery as we deal with the challenge of achieving an educational assessment program that informs instruction, serving students and teachers, and informs the public, assuring parents and communities that what is happening in our schools is meaningful and responsive to both individual and social needs of our citizens.

## Retrospective History

Like all other large, urban school systems, the School District of Philadelphia has a history of systemic assessment of student performance. Like all other American school systems, urban, suburban, and rural, our history reflects the art and science of assessment embedded in testing. For systemwide evaluation of student academic performance, assessment has been defined by commercially prepared, multiple choice, selected-response tests that give nationally norm-referenced comparisons. For local school and classroom assessment, the assessment has been teacher constructed or textbook related tests that are either multiple choice, selected-response items

similar in construction to the commercially prepared tests or short answer, essay, and constructed-response items developed by each individual teacher. Teacher marks for periodic progress reports and for course completion at the senior high school level have always played a major role in the assessment of student performance.

In the 1960's the systemwide test was the Iowa Test of Basic Skills. In the 1970's and early 80's, the California Achievement Test was used. Since 1984, a customized, commerically prepared test based on the California Test of Basic Skills, Form U, has been used.

This customized test developed by CTB McGraw-Hill, now CTB Macmillan McGraw-Hill, offers two components in a single test: (1) a nationally norm-referenced component built on items drawn from the CTBS-U item pool for Reading and Mathematics and (2) a curriculum-referenced component which drew on items specifically developed to match the content and objectives of the School District's curriculum in Social Studies, Science, Mathematics and Reading. This test has been the mainstay of the School District's academic assessment for students in grades 1 through 8.

The customized test has had a great deal of utility in Philadelphia. Data gained from the test has permitted comparisons on nationally normed items, thereby giving us information that we needed for various eligibility requirements for federal and state supplemental program support. This measurement also has helped us respond to inquiries such as "How are our students doing in comparison to other students across the nation?" The customized test was especially valuable in providing longitudinal data on system performance in our own curriculum and in the national norm-referenced material.

For grades 9 through 12, mid-term and final examinations in English, Biology, Chemistry and Physics, Algebra 1 and 2 and Geometry, American and World History have been developed using calibrated items from the CTB item pool. Where items matching our curriculum did not exist, those items were written by the District's curriculum specialists under the supervision of CTB. In other curricular areas such as Anatomy, Calculus, and Health Education, school based mid-term and final examinations have been used. In all cases, most of the examination content was prepared in the multiple choice, selected-response format. Embedded in the English tests developed by CTB are reading passages that yield national norm-referenced data for use with students who qualify for federal or state supplemental program support.

Assessment for kindergarten students has employed a locally developed instrument, the Developmental Checklist. Teachers observe their children and note the developmental level of various skills, behaviors, and understandings. At the close of kindergarten, all students take the California Test of Basic Skills, Form U, which contains a Reading, English Language Arts section, and a Mathematics concepts and applications section. Again, the test yields national norm-referenced data that is used to establish eligibility for supplemental program support and curriculum-referenced data that tells us how well students have learned what we say we teach.

Because the customized tests are designed to yield curriculum referenced assessment, testing is done on a semi-annual basis. Exams are held in mid-January and early June. Schools receive individual student profiles, class profiles, and school profiles. The national norm-referenced data is reported

by percentile rank while the curriculum-referenced data is reported in terms of percent of items correct.

Well in advance of each test administration, teachers receive the learning objectives that will be assessed by the test. After the administration of the test, schools receive their results within four to six weeks. The midyear results become major data sources for both monitoring the local school effort and for school improvement program planning. In March of each year, schools begin to develop their School Improvement Plan for the following year. The end-of-year test results not only serve a monitoring function but also serve as a basis for initial conversations among teachers as they begin the new school year.

Having heard this retrospective history, one may conclude that a lot of testing is done in Philadelphia schools. In addition to the exams noted above, the Philadelphia school district also participates in state-wide testing programs as well as the National Assessment for Educational Progress (NAEP). Some might say too much testing occurs in our schools. We would respond both "yes" and "no." Semi-annual testing does appear to be a bit much in today's environment, but it should be remembered that the focus of the 1980's was the individual child's achievement and the identification of children for supplemental program support.

Prospective History

With the end of the 1980's and the beginning of the 1990's, the focus of assessment in the United States and in Philadelphia has changed from the child to the school. The question now being raised is not so much "What does the child know?" but rather "How is the child being taught and how is the

school preparing children for life in the 21st century?" Business is asking, "How are schools shaping the labor force that we need to keep us and our nation strong, viable economic entities?" Such questions link in a very definitive way educational assessment and instruction. They have led us to reexamine how we assess and what we assess. As a result, we have shifted our attention from decontextualized testing, the general characteristic of standardized tests, to contextualized testing, the characteristic of what is being called "authentic assessment."

Authentic assessment is also known as performance-based assessment. Students are expected to perform specific tasks to solve problems and create solutions, thereby demonstrating their understanding and their ability to use knowledge, skills and behaviors in meaningful, productive, and inventive ways. Instead of being in a selected-response mode (choose the answer that best fits the question), students are being asked to function in a constructed-response mode (tell me or show me your answer or solution and how you reached that conclusion). This "tell me" or "show me" requirement means that teachers must foster ways of thinking and behaving that will enable students to give constructed responses.

Educational assessment in Philadelphia is very much a "work in progress" because we are moving from an old focus to a new focus. For us, the school is the locus of change and the focus of assessment. What we are now doing is creating an overall assessment strategy that will concentrate on instruction and yield information for accountability to our students, parents, community, board of education, state, and national authorities.

It is also probable that we will move from semi-annual testing of all students to annual testing for segments of the student population. At this

time, the major barrier to school-focused assessment, with some form of sampling format, is the requirement to provide individual student data annually, in a pre-and-post test configuration, to satisfy both eligibility and achievement requirements for Chapter 1 program support. In a large urban school district like ours, federal funds are critical program supports.

Because we do not believe that we will ever be free of the need to have norm-referenced data, the Philadelphia school district will continue to employ standardized testing. But we will reduce our reliance on these exams, using them mainly for sampling, so as to get a performance profile for the school. No longer will schools have individual student profiles of standardized test performance. Individual student performance will be shown through other means.

The second segment of our "work in progress" is the development of a variety of performance-based measures that will meet our instructional goals and remain in consonance with the national move to the "three P's," namely, (1) performance tasks, (2) projects, and (3) portfolios. In the School District of Philadelphia we have been encouraging teachers to use these assessment strategies in their classroom testing. To some degree we have been successful, but use of the three P's is still infrequent. The challenge that faces us now is broadening the use of the three P's and extending that use in a systemic fashion.

As a consensus develops on the utilization of performance-based assessment, questions about standards and criteria necessarily arise. The same issues of reliability, validity, and generalization that have been part of the psychometrics of standardized test development prevail in the performance-based assessment development. How do we know what constitutes

"the Good?" What are the common operational definitions for what constitutes portfolios? We have some experience with writing portfolios, but we do not have sufficient experience with standardizing how teachers evaluate them. Reseachers have been asking, "How do we judge the work? How do we communicate with other teachers about the work in the portfolios?"

What are reasonable expectations for student achievement whether it be in portfolios, performance tasks, or projects? How do we communicate to students, parents, teachers, communities, higher education, and business what the performance means? We are a nation that for more than three generations has been conditioned to see and judge by a single score. An example is the SAT score for college admission. Performance-based assessment will not always admit the identification and reporting of a single score. Our lexicon will have to change. Can people accept that change?

The 1915 Army Alpha Test that produced the multiple choice testing model in our country also produced a fast, inexpensive, reasonably efficient test. Performance-based assessment is just the opposite. Performance-based assessment takes time, is costly, and will require teacher reeducation. Teachers will have to learn how to assess and, more importantly, how to teach in ways that demand from students greater discourse, more invention, and greater patience with the messiness of problem solving.

Teachers are the key players in performance-based assessment; we are in the process of building on the interest and expertise that we have scattered throughout the Philadelphia schools. We have invited schools at all levels (elementary, middle, and senior high school) to respond to a Request for Proposals to develop and field test one or more of the three P's. This request provides for a three-year grant, renewable on an annual basis, that

will at the end of the three year period, yield product and process, standards and criteria for assessing student academic achievement in a performance-based assessment mode.

Teachers, principals, parents, and school staff who choose to be part of this research and development effort will have consultant support and whatever professional training they deem necessary. The technical resources of the Office of Assessment, will be aligned with schools that take part in this effort in order to help with the measurement issues that are critical to our ability to use these new forms of assessment in a systemic fashion.

## Conclusion

As we engage in this enterprise of reshaping our educational assessment program for improved instruction, we invite many players to join us, not the least of which are business and higher education. If we are to shape the work force of the future, whether this work force be in the professions, in business and industry, in the arts, or in the public sector, we need to collaborate now.

We are urban educators. We cannot lose sight of our environment. We cannot abandon our cities. We cannot disable public education in our cities through meaningless reform rhetoric. We need to replace rhetoric with reason, resolve, and resources. Performance-based assessment is a step on the road to qualitative change in Philadelphia's public schools.

# ASSESSMENT IN MATHEMATICS INSTRUCTION

Thomas J. Cooney
Professor, Mathematics Education
University of Georgia

Over the past decade assessment has played a particularly important role in shaping public policy toward education and providing a forum from which reform has emerged. On various international assessments our students have fared poorly. The title of the landmark study, The Underachieving Curriculum by Curtis McKnight, et al., which details both achievement and the nature of the curriculum in the United States, was not chosen randomly; it reflects the unmistakable conclusion that the eighth grade curriculum in our schools with its overemphasis on arithmetic and computation is substantially outmoded. In contrast, countries such as France and Japan have eighth grade curricula that focus heavily on geometry and algebra, respectively. The National Assessment of Educational Progress (NAEP) has also demonstrated the difficulty many of our students have in solving problems.

Assessment has laid bare many of our educational shortcomings. As such, assessment has served as an instrument for change. On the other hand, assessments that lead to educational policies based on comparisons of school districts or comparisons among states or to the suggestion that teachers salaries should be tied to student performance on standardized tests have often limited progress in realizing reform. In these and in other instances assessment can be seen as an impediment to change. In considering the role of assessment in realizing reform, I focus here on the question of whether we

are measuring the kinds of outcomes that we believe are important for a technologically oriented society.

## Defining What It Is That We Assess

At the crux of the matter of deciding the form of assesment is the concept of what constitutes mathematics. If we see mathematics as a subject consisting of predetermined formulas that are memorized and applied in rather routine contexts, then assessment is straightforward. We determine whether students can state the formula or use the formula to generate a specific number. The issue here is not philosophical but rather practical and has a great deal to do with what we assess and how we evaluate students' understanding of mathematics. The view of mathematics set forth in Curriculum and Evaluation Standards for School Mathematics (1990), the recent publication by the National Council of Teachers of Mathematics (NCTM), emphasizes much more than the production of single number responses; it requires communication, reasoning, and problem solving.

This view of mathematics is consistent with the publication of the United States Department of Labor, What Work Requires of Schools: "We believe ... that the most effective way of learning skills is 'in context,' placing learning objectives within a real environment rather than insisting that students first learn in the abstract what they will be expected to apply." But this view of mathematics is not always reflected in tests used at the national, state, or classroom level. Multiple choice tests by design cannot assess students' ability to structure real-world problems. In the main, they provide only a limited snapshot of what it means to do mathematics. Consequently, we must raise the question of whether or not

such forms of testing can promote the kinds of mathematical outcomes called for by various academic organizations like the National Council of Teachers of Mathematics.

The notion that standardized tests drive much of what teachers do is legendary. If teachers perceive, as I believe many do, that the best preparation they can give their students for taking such tests is to enhance their ability to compute, then the nature of their classroom activities is predictable. Further, external tests send the message that mathematics is an individual and competitive activity as opposed to a group and cooperative activity designed to solve real world problems. In the Department of Labor publication cited earlier and in the NCTM Standards there is a clear message that cooperation to solve problems is a critical educational objective. But this goal is devalued when the emphasis, real or perceived, is limited to the production of numerical answers.

To illustrate the problem, consider the following items that teachers created to test a "deep and thorough understanding of mathematics" in middle school mathematics:

> 4 1/3 + 2 2/5
> Solve for x: 6x - 2(x + 3) = 1/2x - 10
> How much carpet would it take to cover a floor that is 12.5 ft. by 16.2 ft.?

These items, while sometimes difficult for students, fail to emphasize communication and reasoning beyond the recall of a specific algorithm or a student's ability to think conceptually. In contrast, consider the following item that was developed by the Massachusetts Educational Assessment Program:

> In a recent survey, Americans were asked about ownership of firearms. The findings of the study were that
> -25 percent of American families have at least one handgun;
> -25 percent of American families have at least one rifle; and
> -10 percent of American families have at least one automatic rifle.
>
> A reporter used the following headline on an article she wrote about the study: MAJORITY OF AMERICAN FAMILIES OWN FIREARMS. Should the reporter's editor accept the headline as it is? Why or why not?

Clearly this exercise taps a conceptual understanding of probability and requires the demonstration of an ability to communicate effectively. Unfortunately, only 25 percent of the 12th graders surveyed said "no" and provided a correct reason (e.g., families may own more than one kind of gun). Thirty four percent said "no," but provided incorrect reasons (e.g., 25 percent is not a majority). Thirty five percent said "yes." Eight percent provided other incorrect answers (e.g., the headline is not catchy enough).

We live in an era of accountability and efficiency, and there is much that is good about such demands. But the fallout for education is often a focus on the easy part of assessment: determining whether or not students can do mechanistic and predetermined exercises as quickly as possible. Teachers get the message that their students should be able to perform the "basics" effectively and efficiently. Indeed, some research reports emphasize that the "expert" teacher is one who can cover more problems than the "novice" teacher, again placing an emphasis on efficiency. But efficiency has its limitations; it does not lend itself to understanding conceptually important ideas like those reflected in the "firearm" problem.

To illustrate further, I would like to share a result of an evaluation study that I conducted in an elementary school. Over a two day period I observed every classroom in which mathematics was taught. There was much to

be applauded. The lessons were well organized and the teachers and students were busy doing mathematics. But there was one disturbing observation that has stayed with me: on only two occasions did I witness a question to which the students did not immediately know the answer. Perhaps what I observed was atypical. But if it was typical (observations in other schools suggest so), consider the consequence of students studying mathematics over a period of years in which they think mathematics is primarily a matter of being able to produce a number with great proficiency.

I contend that when students finally encounter situations that do require the demonstration of communication and reasoning, as in the "firearm" problem, they will realize there is something wrong with the way their instructor is teaching mathematics, there is something wrong with the mathematics, or, perhaps of more serious consequence, they will conclude that they are no longer any good at doing mathematics. In contrast, consider the following scenario described to me by a Japanese colleague, who claimed the following type of problem was used in a Japanese elementary classroom over several class periods.

> Jack is in a race with 15 other boys. At the beginning of the race Jack is fifth from last. At the end of the race he is third. How many boys did he pass during the race?

Once solved, general solution strategies are examined that lead students to work out the following, more complex problem.

> Jack is in a race with 200 other boys. At the beginning of the race Jack is 50th from last. At the end of the race he is 15th. How many boys did he pass during the race?

I mention these problems because they convey an appreciation of

mathematics that differs from one in which mathematics is seen as a subject consisting only of rules and formulas to be memorized -- a perception shared by the majority of eighth grade students in the United States. This limited perception must change if reform is to become a reality and not an illusion.

Let us be clear: assessment conveys a strong message about what is valued. When a teacher tells students what will be on Friday's test, she is conveying what aspect of mathematics she values and consequently what the students will value. Similarly, if our state and national tests do not reflect the sense that mathematics is a language we use to communicate and reason, then we can hardly expect teachers to "buck" the system as they structure mathematical activities for their students. To repeat, the crux of the matter is what we consider mathematics to be. From this foundation we can address issues related to assessing students' understanding of mathematics.

## Some Issues in Assessing Mathematical Understanding

There are various issues that remain when we consider assessing performance on more authentic items. One has to do with scoring. Certainly, it is much easier to test whether students can find the arithmetic mean of five numbers than it is to measure their interpretation of a headline that might read, "AVERAGE TEACHER MAKES $28,000." Objectivity is another concern. Certainly objectivity is a vital issue if the results are to direct policy matters. A further issue is time; it takes much longer to evaluate responses involving communication and reasoning than it does to determine whether the student has generated the correct number. Time costs money. In the early days of NAEP, open-ended items were used, but this approach was abandoned

because of difficulties in scoring and the time and expense of evaluating the responses to these complicated mathematical problems. Still another issue is what form assessment should take given the ready availability of calculators and computers. Although we once could justify asking students to multiply decimals, such an assessment item seems out of place in today's world.

If we are to take seriously the concept of mathematics as a language through which we communicate, reason, and solve problems, then our assessments must reflect such outcomes. The use of performance items in which students are required to construct arguments and demonstrate their reasoning is gaining hold in some states (e.g., North Carolina, Massachusetts, and California), and it is being considered as part of a national assessment program as well. At present, however, the use of such performance items on state or national tests is limited, although the tendency is to move away from a total reliance on multiple choice tests. At the classroom level, teachers are reluctant to use performance items because of the aforementioned issues and a lack of problems or tasks that they can readily utilize. This reluctance raises an issue related to teacher education. Almost without exception the teachers I have met are concerned with what their students are learning. But often they have a limited conception of what it means to do mathematics, thus raising a serious question of how we can counter this limitation in our teacher education programs.

We must realize that we have several challenges facing us. If we are to help students develop the kind of expertise that business and industry wants, we must nurture students' ability to think and reason mathematically. We

cannot expect this to occur when assessment fails to take into account a broader view of mathematics. While computational skills and solving problems involving the production of a single number may be a part of the picture, they are not the whole picture -- particularly given the availability of inexpensive calculators. Long ago Alfred North Whitehead in his classical work The Aims of Education warned against teaching mathematics so that students perform "intellectual minuets" in the absence of understanding the holistic nature of mathematics. Tests and examinations of all forms must move away from assessing "intellectual minuets" and toward significant mathematical outcomes. Otherwise, they are certain to become a formidable impediment to reform.

# ASSESSMENT IN SCIENCE INSTRUCTION

Audrey B. Champagne
Professor of Education & Chemistry
University at Albany, SUNY

In discussing assessment of science instruction, one can best begin by asking two major questions:

What is the problem?

School science tests are dominated by short answer items and predictable end-of-chapter quantitative exercises euphemistically called problems. Objective tests do not reflect the nature of science or the valued goals of school science.

Experts agree that performance on objective tests is a weak indicator of scientific competence. Performance of students engaging in the solution of challenging problems and inquiring into the nature of the physical world is a much stronger indicator.

However, the defenders of objective testing remind us that factual information and computational skills are necessary component skills of scientific problem solving and inquiry competence.

What will it take to improve the situation?

To turn around the existing situation will require (a) the development of assessment exercises that more accurately reflect the nature of science and (b) educating and encouraging science teachers to use the new approaches to science-testing effectively.

Testing is a pervasive aspect of schooling that serves many purposes, both educational and political, such as rating students, rating the effectiveness of instructional programs, and determining the impact of public resources appropriated to education. For educators, the use of information from tests to make instructional decisions is most important. The vast majority of these decisions are based on information from objective test instruments, paper and pencil examinations that require the student to choose

a correct answer from several possibilities. Multiple choice and matching are common formats. Objective tests are based on the assumption that knowledge is an accumulation of unordered bits of information and that for every question there is one right answer. As objective tests have come to play a larger role in American schools, increasing numbers of teachers and administrators have become concerned about their negative effects. Originally tests were assumed to be neutral evaluating instruments, like thermometers, which could be dipped into a school and taken out without having any further influence on the institution. Regrettably, the impact of objective tests expands far beyond attempting to measure what students learn. Teachers, motivated by a desire to have their students perform well and knowing that their teaching ability will be judged by their students' scores, often teach to tests. Thus, testing modes have a profound influence on teaching, often limiting classroom activities to exercises that will be on the test.

Objective examinations evaluate the extent and quality of the students' information base and their ability to solve academic problems. But most experts in the field agree that these are weak indicators of other aspects of scientific literacy. Objective tests are educationally limiting because these exercises cannot assess students' ability to solve open-ended problems or to engage in scientific inquiry. As a consequence, such critical abilities are given short shrift during instruction.

Objective instruments are not without virtue. They are inexpensive and easy to administer to large numbers of students. Standardized tests are an efficient means for rank-ordering large numbers of individuals. Such tests are easy to score. Since they are standardized, the scores can readily be

used to make comparisons across localities -- cities, states, nations. And they fit the American cultural assumption that there is a single dimension of ability distinguishing winners and losers.

## Alternative Assessments

Alternative assessment is based on the assumption that effective learners organize their knowledge into structured patterns. These patterns tie bits of information together so that they will be remembered. Subtle evaluation methods must be devised to capture these memory structures as well as the reasoning processes that are the basis of scientific literacy.

Performance exercises and portfolios are two strategies of alternative assessment. Performance exercises ask the student to do something concrete so that an observer can evaluate the student's knowledge, mental processess, problem solving, and inquiry capability. Portfolios are collections of student products that allow assessment of a student's attainments.

Performance assessments can be divided into two categories. Academic performance tasks require students to show their skills in laboratory practical exams and other closed-ended school problems. For example, the ability to use common laboratory equipment and to design and conduct simple experiments can be measured. Authentic assignments involve real-world, open-ended tasks such as deciding on the best place to locate a landfill or recommending the best brand of paper towel to the school janitor.

A subset of performance testing is "dynamic assessment" or "developmental assessment." In this case, the goal is not to measure what the student already knows. This mode seeks to analyze the child's potential for intellectual growth. During the task presentation, the evaluator can

provide feedback, evaluation and support as needed. Measurement then focuses on how effectively the student responds to this feedback in constructing new learning.

Performance assessments are often administered to groups of three or four students for two reasons. Authentic tasks are often complicated. It is unlikely that one student alone has the skills, time and motivation to do them well. Second, group work is becoming increasingly important in the science curriculum. Professional scientists almost always work collaboratively; this is an important experience to impart to the next generation.

Portfolios are collections of materials that provide evidence of a student's achievement. Over a period of time, teachers and students work together to determine which lab reports, journal entries, research papers, and other educational exercises should be kept as evidence of progress. These products reflect each student's level of competence over a period of time. This mode of assessment also encourages students and teachers to reflect on the items to be placed in the portfolio. Consequently, it helps students to develop self-assessment strategies that will be useful in all domains. Finally, portfolios are sensitive to context. Different age levels and different sciences will elicit a range of choices about what to enter into the portfolio.

Proponents of alternate assessments claim they have other virtues. Alternate assessments enable the evaluation of a wide variety of valued outcomes. Tests which focus on the wide-ranging skills of science literacy will motivate instructional time spent developing these skills. Alternate instruments also provide much more information about what each student knows

and is able to do. Laboratory skills, the ability to work on open-ended problems, and the ability to think collaboratively in groups are all promoted through alternate assessments. It is also possible to get a picture of an individual's intellectual development over the course of a project, since time constraints are fewer than in conventional testing.

Alternate assessments modify the distinction between instruction and evaluation, reducing pressure on students to perform well under testing conditions. They also offer greater opportunities for special education students and students with inadequate English skills to demonstrate competence in their own way. In our society, for example, girls are often taught not to compete. Tests which place a positive value on cooperation may be more accurate measures of girls' skills.

In deciding which of these formats to use, the evaluator must take into consideration the purposes of the assessment. Tests serve internal purposes of classrooms and local districts or external purposes of monitoring progress in states or across the nation. In a classroom, a teacher can choose formative evaluation which helps inform future instruction or summative evaluation which shows how much the students have learned. Alternative exercises can help a student to learn more or can be used to demonstrate student competence. Performances and portfolios can be designed either for rating individual progress or for tracking large numbers of students to inform policy decisions. Dynamic assessment is best for monitoring individual progress.

The major weakness of alternate modes is that they exist mostly in the minds of their advocates. With the exception of a few localities and states, such as New York, California, and Connecticut, alternate assessments have not

been widely used. Thus, there is little experience with ways of making them live up to their promise. As more schools and districts institute alternate forms, research is needed to identify weaknesses and devise means to circumvent them.

## Policy Issues

Pedagogical, technical, and social/political issues abound in testing. They interact in very complex ways. One example is the interplay of testing and the underrepresentation of women and minorities in science.

The concern for women and minorities in science cuts across political, social, pedagogical and technical concerns. It is a political issue because of the need to have an adequately trained workforce able to deal with the technology-based economy of the next century. Also, our democracy depends on a scientifically literate voting public. It is a social issue because of concerns for equality and justice in our society.

Science education for females and minorities is a pedagogical and psychological issue as well. Many questions on current standardized tests have been shown to be prejudicial against females. Research indicating that from second through twelfth grade, boys are more enthusiastic about competition than girls has serious implications for competitive individualistic testing. If the underrepresented students work more effectively in non-competitive situations, then the best way to develop and truly assess their knowledge and skills is in groups.

Technical issues are also raised by a consideration of the academic needs of others. Disabled students, speakers of non-standard English, and special-education students all have unusual needs that may be differentially

served by alternate testing formats. The inherent flexibility offered by portfolios and performance testing is surely an advantage. Without the pressure of limited time, these students have a better opportunity to select their best work or to continue to revise and improve their work before it is evaluated. However, since new formats do not allow students to get by with rote memorization, they will reveal weaknesses in ability to think or communicate. Developers of alternate assessments will need to take into account the special needs of these populations. To date, almost no research has been done in this area.

Concerns for assessment and the greater inclusion of women, Hispanics, and Blacks crosscut all areas. Each of the issues can be used as a lens to view the others. Pedagogical issues have political overtones, particularly when teacher unions become involved. Legislative initiatives may have technical implications as test makers try to comply with new laws. Technical changes involving new ways to use computers may influence both teaching and politics. Sorting out these issues will keep educational researchers and developers busy well into the next century.

# ASSESSMENT IN READING & WRITING INSTRUCTION

Sheila W. Valencia
Associate Professor of Education
University of Washington

The assessment of reading and writing is a priority across the United States. Simply in terms of numbers, we can document its prevalence: 45 of the 50 states regularly conduct statewide assessment of reading and 33 of 50 conduct writing assessment; the National Assessment of Education Progress (NAEP) has included both reading and writing since its inception in 1969-70; and every available commercial standardized test battery includes subtests in reading and language arts. Add to these figures the fact that much of our supplemental funding for education is based on students' abilities in reading and writing (e.g., Chapter 1) and that reading and writing learning and assessment is highlighted in the President's America 2000 plan, and the importance of these evaluations is confirmed.

## Setting the Stage for Changes in Reading and Writing Assessment

For years, reading and writing were measured by the ubiquitous multiple-choice test, but in the late 1970s and early 1980s assessment began to change. Two factors converged to produce this change. First, educators began to react to the pressure of "high-stakes" tests brought on by the accountability movement. Indeed, the potential consequences of poor scores had a narrowing and negative influence on curriculum and instruction, causing teachers to believe they had to limit their teaching to the content and the format of the test. Second, 20 years of research in cognitive psychology and instructional strategies found its way into many classrooms. In essence, we

came to understand and focus on the processes of reading and writing rather than just examining their products or the end results of learning. We moved away from a compartmentalized, skills-based view of learning to read and write toward an interactive, contextualized, holistic view. The result was new curricula and new teaching strategies that emphasized in-depth reading and writing of more genuine, "authentic" texts, activities that were very different than those tested on most literacy assessments.

The misalignments between what was happening in classrooms and what was being assessed on most standardized tests became obvious. For example, students were reading full-length books, short stories, and trade books in class but they were encountering short paragraphs of 100-150 words on reading tests. In class, students were writing papers that required days or weeks to complete, analyzing the merits of particular characters, analyzing authors' styles, evaluating and making judgments about what they were reading and writing. On reading tests they were answering low-level, literal, multiple-choice questions and on language arts tests they were recognizing errors and then choosing the correct word or sentence from a list. These discrepancies, coupled with the recognition that testing was bound to influence classroom practice, prompted literacy educators to call for changes in assessment that would give priority to understanding and thinking about reading and writing in genuine and important ways.

New reading and writing assessments, like assessment in all other disciplines, serve three primary purposes: accountability, improvement of instruction, and improvement of learning. Each of these purposes has a different primary audience -- policy-makers and administrators (accountability), teachers (instruction), and students (learning).

Furthermore, each purpose may be met by assessment controlled at the national, state, district, and classroom levels. The issues of purpose, audience, and control are important considerations in new reading and writing assessments.

Regardless of the purpose or locus of control, all new assessment efforts stress authenticity. Authentic assessment tasks engage students in reading and writing that are representative of important, realistic tasks they will meet in their lives. At a minimum, authenticity is manifested in reading and writing assessments by the use of a variety of longer, more natural reading passages, more inferential and critical reading questions, or a combination of multiple-choice and open-ended questions where students provide written responses.

Authenticity is carried a step further in new performance-based assessments that require students to perform a specific task such as writing about a frightening experience, performing a musical composition, or conducting a science experiment. The point is for students to engage in a complete, complex activity that requires them to use their knowledge to generate answers or products, rather than recognize correct answers from a list of wrong ones. Currently, the most common performance-based assessment is the direct writing assessment in which students are given 15-90 minutes to write a paper in response to a given prompt. Performance-based assessments in reading are much less common at this time, but when they do appear, they often consist of students writing responses to questions.

Finally, many reading and writing educators are promoting portfolios as the most authentic of the new assessment alternatives. Portfolios are on-going collections of a student's work that include samples carefully selected by the teacher and/or by the student to represent the student's

capabilities. The contents of the portfolio are collected as part of the regular instructional program, and because there are several entries, they provide direct evidence of the student's progress over time. In most cases, students also annotate some of the entries and reflect on their own work, giving them an active role in their own evaluation and learning.

## New Assessments

Large-scale assessment. Several new assessments have been developed for large groups of students at the national or state level to address accountability issues. One example is the 1992 National Assessment of Educational Progress in reading, which represents a move toward authenticity while still maintaining a format that can be used with large numbers of students. The test requires students to read "longer, naturally occurring stories and articles, as well as documents" and to answer questions that emphasize critical thinking, interpretation, and personal response. Approximately 50 percent of the questions require students to generate their own written responses. A separate NAEP writing assessment requires students to write for approximately 20-40 minutes in response to a given topic. Students are randomly assigned to one of several kinds of writing tasks so that NAEP collects data on the different kinds of writing that students will be expected to do.

At the state level, more than 60 percent of the states continue to use one of the commercially available, multiple-choice, standardized tests in reading and language arts. Although many of these tests have attempted to address the authenticity issue through the use of better reading passages and questions, most still resemble the problematic format and content of the

past. A few publishers, however, have begun to develop tests requiring more open-ended written responses, and several offer the option of having students participate in direct writing exercises. A recent survey of state assessment practices indicated that 31 states now use direct writing assessment but that only 7 have made any attempt at more authentic, performance-based reading assessments, and most of these are simply multiple-choice tests enhanced with open-ended questions.

Although writing is tangible and relatively easy to collect, the reading process remains hidden much of the time, making its evaluation difficult and elusive. Some states, however, have made efforts to rethink reading assessment and have developed their own reading and writing assessments or have worked with publishers to develop customized tests. Several of these instruments include a variety of longer passages and higher level comprehension questions. A few have included assessments of other important reading factors, such as students' background knowledge, reading habits, and reading strategies; and a few require students to write answers to open-ended questions.

Several states are experimenting with more performance-based literacy measures (e.g., California, Kentucky) in which all students must perform a particular task such as conducting research and writing a report on heart disease. Two states, Maryland and Vermont, have just implemented very innovative assessements. The Maryland School Performance Assessment Program includes only open-ended items and integrates reading and writing assessment. This multi-day assessment requires students to complete in-depth tasks around a piece of literature and to produce a written product. Vermont's Writing Assessment Program uses writing collected in a portfolio and a uniform 90

minute assignment to assess students. Statewide scores are based on the evaluation of the portfolio as a whole, a "best piece" each student identifies in his portfolio, and the uniform writing assignment.

Classroom assessment. Although large group accountability-focused assessments still dominate literacy assessment, there is an interest in fostering improvement of teaching and learning by situating more assessment directly in the classroom. Current thinking suggests that because a major objective of assessment is the improvement of teaching and learning, assessment must be on-going, teachers should have a direct role in the evaluation of students' abilities, evidence must be readily available to teachers and students so they can make sound decisions, and students should have an active role in the learning and evaluation process. Tyler & White go further and suggest that purposes of accountability would be best served by looking for information in the classroom, where we are able to find the most direct and most authentic evidence of learning. They note that assessment should "come out of the classroom process, not be imposed on it like a foreign body." Even new efforts to build a national examination system, such as the New Standards Project, call for the "three P's -- timed Performance examinations, student Projects (group and individual), and Portfolios of student work"; all of which acknowledge the importance of classroom-based assessment.

Portfolio assessment is the most commonly discussed means of collecting on-going, authentic evidence of students' work that involves both students and teachers. Although only three states report any attempts to develop and implement portfolio assessment, evidence from the professional literature suggests that large numbers of teachers, schools, and school districts are

experimenting with portfolios. Most are convinced that portfolios foster sound assessment because they provide direct evidence of the students' progress over time and involve students in self-evaluation. Educators also feel that portfolios improve the quality of teaching and learning and raise the level of communication with parents. Informal evidence from my work and that of others suggests that this is so. Nevertheless, at this point, portfolios are still an amorphous concept. There are many different ways to define portfolios and to determine what goes into them, who will select work for inclusion, how they will be evaluated, and who will get the information. Clearly, portfolios deserve attention, but they are still embryonic as assessment tools.

## Issues in Reading and Writing Assessment

Howard Gardner reminds us that it has taken a century to develop mediocre standardized tests and that it would be utopian to expect that we can take new assessments and "whip them into shape right away." New assessments will take time and they will require new ways of thinking about the unique issues they present. These issues can be conceptualized as issues of content, implementation and use, and measurement characteristics. They provide a sampling of questions that must be addressed as new reading and writing assessments are developed.

### Content:

-Who should decide and how should we decide what to assess?

-We have little insight into authentic assessments of reading. What would authentic reading assessment look like? How can we make reading "visible" for assessment purposes without disrupting or changing the reading process?

-Do we introduce bias and contamination into reading assessment when most new assessments require students to write answers? Are there other response modes we should explore?

-How much do classroom portfolios reflect the classroom context and how much do they reflect the abilities of individual students? For example, if students are presented with very different instructional activities how does that influence the composition of their portfolios? How much consistency do we need for portfolios to serve as assessment tools?

Implementation and Use:

-How do we build sound classroom assessment of literacy given the limited knowledge base and training of most teachers?

-How do we keep these new performance-based assessments, with their very specific tasks and projects, from becoming corrupted in the way assessments of the past have been corrupted?

-What can we conclude about students' overall abilities in reading and writing from their performance on specialized tasks found in performance-based assessments? How many samples of a student's work do we need to be sure we are making the best assessment of the individual or of the group?

-What will be the impact of these new assessments on teaching and learning?

-How do we keep the focus on enhancing learning and instruction while still meeting the needs of accountability? How can we fulfill the different purposes of assessment?

Measurement Characteristics:

-How do we assure the reliability and validity of classroom-based and performance-based assessments?

-What do we know about the cultural bias of these new assessments? Are there other types of bias that surface as assessments are embedded in the classroom and assessment tasks are made specific?

## Recommendations

I have described the past and present of reading and writing assessment. That reading and writing assessments are important for our society and that they are changing is obvious. Less obvious, however, is what the future holds. There is little doubt in my mind that these changes are crucial and

that we must encourage all efforts to improve literacy assessment. Our educational system and our society require assessments that represent high quality reading and writing and that promote the best instruction and learning we have to offer.

Clearly, there are many unresolved and important issues that must be studied. This is not to suggest that we stop our efforts but rather that we remain committed to monitoring new assessment as we watch them implemented throughout the country. We can improve new assessment efforts in several ways:

First, we must support collaborative efforts among schools, states, federal agencies, communities, businesses, and universities to develop and implement reading and writing assessments. Developing new assessments is a time-consuming, rigorous, and expensive endeavor that requires expertise and commitment on all fronts.

Second, we must support new assessments that are tied to curricular and instructional reform. New assessment, whether large-scale or classroom-based, must begin with knowledge of what is important to assess, not with what is easy or convenient to measure. In addition, sound assessment requires knowledgeable teachers, and this requirement demands substantial staff development and support. Without a combined effort to work on curriculum, instruction, and assessment, new assessments are likely to fail.

Third, we need considerable research to develop and evaluate new assessments. We need to investigate and document their measurement qualities and their intended and unintended impact on education as they are being tried in schools. We need to be thoughtful about new possibilities.

Finally, we must continue to be sensitive to different purposes and

audiences so that we can provide the kinds of information needed and, at the same time, continue to promote improved teaching and learning.

These are difficult challenges, but the success of these profound changes in assessment is vital to the reform of education in the United States.

# SUMMATION

Robert Rothman
Associate Editor
Education Week

The recent debates about assessment remind me of Woody Allen's story about the two women who have just returned from a Catskills resort. One says, "The food at that resort was terrible." And the other replies, "Yes. And such small portions!"

In testing, we hear that tests often do not serve learning, that they narrow the curriculum, and that they force teachers to focus on basic skills as opposed to higher-order abilities. But we also hear that we need more tests; we need national tests.

What the essays collected here add in loud and clear tones is that there is a widespread movement to develop new and better ways of measuring student learning. This movement encompasses all the players that we have heard from over the past two days: government, where the Bush Administration and the nation's governors are looking for ways to measure progress on the national education goals; business, where the Business Roundtable has pledged to judge schools' success or failure on the basis of student outcomes; school administrators, who are looking to hold schools accountable for meeting curricular goals and who want to target resources where they are most needed; and teachers themselves, who want to be sure assessments foster the type of instruction that can help students learn.

This national concern, coalescing in a national assessment movement, has the potential to lead to a dramatically changed education system over the

next few years, a "world-class system," as Pat Forgione describes this initiative.

But there are many questions that must be answered first and many pitfalls that must be overcome. As Audrey Champagne observes, these new forms of assessment "exist mostly in the minds of their advocates," and where they have been tried, it has never been on the scale that is envisioned today.

One of the key questions people raise about assessment is its purpose. We have heard many purposes stated here over the past two days, and to some extent, they must be reconciled. As Stanley Bernknopf emphasizes, tests cannot serve two masters.

Stanley Bernknopf tells us that Georgia's assessment system is designed to monitor whether schools and school districts are meeting state curricular objectives. These objectives are designed to allow comparisons with national performance and ensure that graduating seniors meet established state goals. To make these assessments, the state has chosen a very practical solution, matrix sampling, which reduces costs and burdens on instructional time. But while that solution provides the Georgia Department of Education with information on schools and school districts, it does not generate information on individual students, which Pat Forgione says would help parents see if their children are meeting established standards of performance. And the sampling technique does not necessarily encourage every classroom to switch to the type of instruction the new assessments measure, a critical step if schools are, in Thomas Cooney's words, to move from "current city to vision village." If a specific classroom is not part of the sample, would its teacher have any incentive to teach students to communicate mathematically,

to explain what a calorie is used for, or to develop the reading and writing competencies that Sheila Valencia speaks of?

Philadelphia, on the other hand, is trying the opposite tack: sampling on its standardized tests and performance assessments at the classroom level. But this approach raises other issues, as Rita Altman points out. How do you get teachers to agree on evaluations of performance and what the "good" is?

Educating teachers to be able to answer those questions, as well as to teach in the ways those new assessments value, is a major task. It will take money. If those are the kinds of assessments schools need, then everyone in government, business, and education will have to muster what Stephen Aiello calls the "political will" to make them come about.

But it will also take more to make the assessment movement successful. There has been a belief -- a hope, really -- that somehow these new assessments would wipe out the gaps in achievement between advantaged and disadvantaged students. There is some evidence to support that view. There is also evidence, from the National Assessment of Educational Progress and other sources, that, absent other changes, performance-based assessments might widen the gaps. Disadvantaged students are less likely now to have received the kind of instruction to solve communication and writing problems that would enable them to do well. Giving those students a fair shake, particularly if the results are to be used to hold them or their schools accountable, will take major effort. It will also take more than changes in schooling. Disadvantage starts before children walk in the schoolhouse door.

There are many other questions. Archie Lapointe speaks of the importance of standards. There is probably more agreement than disagreement on what

students should know and be able to do, although there will be considerable squabbling over the details. But there is also some experience to build on. As Rita Altman says, portfolios are not new to artists. And Sheila Valencia has described the growing use of direct-writing assessment.

Every participant at this conference wants to share Archie Lapointe's optimism. The assessment movement raises as many questions as it answers. No one can predict with certainty the most lasting form and value that new assessments will contribute to educational reform.